



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Inter-Observer Variability of the Apgar Score of Preterm Infants between Neonatologists, Obstetricians and Midwives

Arri, Sandra Jasminder ; Bucher, Hans Ulrich ; Merlini, Massimo ; Fauchère, Jean-Claude

Abstract: Objective To assess the inter-observer variability of the Apgar Score (AS) across various perinatal Health Care Providers (HCP) taking care of newly born premature infants in the delivery room. **Methods Design:** Prospective observational study. **Setting:** 4 general hospitals and 3 university hospitals in Switzerland. **Subjects:** 43 neonatologists, 68 obstetricians and 55 midwives assessed the AS from 15 video sequences showing delivery room stabilisations or resuscitations of 15 preterm infants born below 34 0/7 weeks gestational age. **Results** Overall and for all observers, the mean inter-observer variability was low (ICC 0.72). There was a significant difference between the professions ($p < 0.001$) and hospitals ($p < 0.001$). The AS assigned by neonatologists for this group of preterm infants were significantly higher than the scores given by midwives ($p = 0.001$). The scores assigned by obstetricians were the lowest for all infants; the difference from neonatologists being -0.53 (pairwise comparison). There was no significant difference between the AS assessed by professionals working in university hospitals compared to HCPs from general hospitals ($p = 0.86$). For all observers and in the majority of the sequences, heart rate showed the lowest and skin colour the highest standard deviation. **Conclusion** Our study revealed a relatively high inter-observer agreement in assessing the AS for premature infants among all perinatal health care professionals for the whole group of infants. A significant difference however was seen between the AS given by the different perinatal professional groups and between hospitals. A clearer definition and assessment method of each Apgar parameter in the setting of infants born premature and of resuscitation measures are needed. This may contribute to reduce the variations between professionals and hospitals, and to increase the value of this scoring within national and international databases to describe study populations for research, for benchmarking in neonatal intensive care and for comparison of outcome data.

DOI: <https://doi.org/10.24966/ncp-878x/100024>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-166233>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 3.0 Unported (CC BY 3.0) License.

Originally published at:

Arri, Sandra Jasminder; Bucher, Hans Ulrich; Merlini, Massimo; Fauchère, Jean-Claude (2018). Inter-Observer Variability of the Apgar Score of Preterm Infants between Neonatologists, Obstetricians and Midwives. *Journal of Neonatology and Clinical Pediatrics*, 5:24.
DOI: <https://doi.org/10.24966/ncp-878x/100024>

Research Article

Inter-Observer Variability of the Apgar Score of Preterm Infants between Neonatologists, Obstetricians and Midwives

Sandra Jasminder Arri¹, Hans Ulrich Bucher¹, Massimo Merlini^{1,2} and Jean-Claude Fauchère^{1*}

¹Division of Neonatology, University Hospital Zurich, Zurich, Switzerland

²Division of Biostatistics, University of Zurich, Switzerland

Abstract

Objective

To assess the inter-observer variability of the Apgar Score (AS) across various perinatal Health Care Providers (HCP) taking care of newly born premature infants in the delivery room.

Methods

Design: Prospective observational study.

Setting: 4 general hospitals and 3 university hospitals in Switzerland.

Subjects: 43 neonatologists, 68 obstetricians and 55 midwives assessed the AS from 15 video sequences showing delivery room stabilisations or resuscitations of 15 preterm infants born below 34 0/7 weeks gestational age.

Results

Overall and for all observers, the mean inter-observer variability was low (ICC 0.72). There was a significant difference between the professions ($p < 0.001$) and hospitals ($p < 0.001$). The AS assigned by neonatologists for this group of preterm infants were significantly higher than the scores given by midwives ($p = 0.001$). The scores assigned by obstetricians were the lowest for all infants; the difference from neonatologists being -0.53 (pairwise comparison). There was no significant difference between the AS assessed by profession-

als working in university hospitals compared to HCPs from general hospitals ($p = 0.86$). For all observers and in the majority of the sequences, heart rate showed the lowest and skin colour the highest standard deviation.

Conclusion

Our study revealed a relatively high inter-observer agreement in assessing the AS for premature infants among all perinatal health care professionals for the whole group of infants. A significant difference however was seen between the AS given by the different perinatal professional groups and between hospitals. A clearer definition and assessment method of each Apgar parameter in the setting of infants born premature and of resuscitation measures are needed. This may contribute to reduce the variations between professionals and hospitals, and to increase the value of this scoring within national and international databases to describe study populations for research, for benchmarking in neonatal intensive care and for comparison of outcome data.

Keywords: Adaptation; Delivery room; Neonate; Newly born; Resuscitation; Stabilisation

Abbreviations

AS: Apgar Score

NICU: Neonatal Intensive Care Unit

HCP: Health Care Provider

Introduction

The Apgar score, as designed by Virginia Apgar in 1953, was primarily developed to assess the effect of maternal analgesia and anaesthetic during labor and different obstetrics techniques on the immediate neonatal adaptation of infants born at term, and also to guide neonatal resuscitation measures directly after birth [1,2]. From a locally developed clinical assessment tool, it rapidly gained international acceptance as the first standardized method, and eventually as the “gold standard” to evaluate and to document the immediate neonatal adaptation as well as the efficacy of neonatal stabilisation and resuscitation measures in the delivery room [3-7]. Over the past decades, a low AS at 5 minutes of life has also gained interest regarding its prediction of neonatal mortality [8-13] and long-term morbidity [11,13-15].

Due to the advances in neonatal medicine over the last 50 years, an increasing number of very preterm infants are being offered resuscitation measures and intensive care; yet the AS has not been adjusted to this population of immature infants. No consistent data are available on the interpretation and on the applicability of the AS in premature infants. When compared to term infants, preterm infants may well be given lower AS only due to the immaturity itself, even when the immediate adaptation is not impeded by cardio-respiratory problems [16]. This fact questions the prognostic significance of the AS for this population of preterm patients although it has been suggested that low AS may be of predictive value regarding neonatal mortality of infants born premature [9,10,17,18].

***Corresponding author:** Jean-Claude Fauchère, Division of Neonatology, University Hospital Zurich, Zurich, Switzerland, Tel: +41 442553584; E-mail: jean-claude.fauchere@usz.ch

Citation: Arri SJ, Bucher HU, Merlini M, Fauchère JC (2018) Inter-Observer Variability of the Apgar Score of Preterm Infants between Neonatologists, Obstetricians and Midwives. J Neonatol Clin Pediatr 5: 024.

Received: June 26, 2018; **Accepted:** July 24, 2018; **Published:** August 08, 2018

Copyright: © 2018 Arri SJ, et al., This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Livingston and co-workers showed that the inter-observer variability was smaller when assessing the AS in infants born at term than in those born preterm [19]. Elements of the score such as skin colour, muscle tone and reflex irritability very much depend on the maturation, and thus on the gestational age of the newborn infant [20-22].

A further problem in documenting the neonatal adaptation in the delivery room may be due to the fact that many Health Care Providers (HCPs) are not sufficiently trained in assessing the AS which is mirrored by a high inter-observer variation [2,23,24]. Using three written clinical scenarios of two term and one preterm infant, Gupta and co-workers showed that a simple clarification of the AS such as proposed by Lopriore can improve the inter-observer variability between paediatricians, obstetricians, nurse practitioners and neonatology fellows [25,26].

Significant differences in assessing the AS challenge the value of this scoring within national and international databases to describe study populations for research, for benchmarking in neonatal intensive care and for comparison of outcome data. Therefore and using a series of video recordings of the immediate neonatal adaptation, the aims of our study were to explore the inter-observer variability between different perinatal professional groups taking care of newly born infants in the delivery room, namely midwives, obstetricians and neonatologists, with regard to assessing the AS in infants born preterm in a setting as realistic as possible. We also aimed at studying the influence of different hospital settings on the variability. Moreover, we included bigger numbers of participating perinatal professionals to improve the statistic validity.

Methods

Based on the Swiss Human Research Act (Art. 7, category A; Coordination Office for Human Research, Federal Office of Public Health), this non-clinical observational study was exempt from the requirement for approval by the Ethics Committee of the Canton Zurich and by the Clinical Trial Centre of the University Hospital Zurich as no patient data or health-related data of the participating HCPs was assessed. Participation was voluntary, the determination of the Apgar scores was anonymously collected, and no identifying data such as name, gender or age was included. The chosen video sequences were used only with written parental permission. Care was taken to avoid any identifier of the infant and of the attending HCPs. The eyes of the infants, however, were not covered by a black bar in order to allow the assessment of the facial expression of the given infant.

The video sequences were recorded using a professional digital video camera with a spotlight and a microphone (Panasonic DVC-Pro HD P2, Panasonic Corp. Osaka, Japan; Sennheiser Microphone, Sennheiser Electronics, Wedemark-Wennebostel, Germany; Dedolight DLH4, Dedotec Inc. Ashley Falls, MA, USA). The camera was attached to a movable pivot arm mounted on the ceiling above the resuscitation cot in the labour ward of the Perinatal Centre at the University Hospital Zurich. This camera was positioned such to acquire a clear view of the newborn without disturbing the professionals taking care of the neonate.

We enrolled 15 preterm infants with a gestational age below 34 0/7 weeks. These newly born infants were video recorded while receiving various stabilisation measures or resuscitation interventions. No chest compressions and no medications were given. From each of the video

sequences, 15 seconds were extracted on which the following four parameters of the AS were clearly visible, namely respiratory effort, muscle tone, reflex irritability, and skin colour. These sequences were chosen independently from the Apgar assessment time points at 1, 5 or 10 minutes of life. Heart rate was provided visually by finger tipping, no oximetry reading was shown. Audio sound was eliminated in order to avoid a bias through audible AS assessments and comments performed by the attending staff. The infant's crying could be estimated by mimic changes. These video sequences were then shown to midwives, obstetricians and neonatologists regularly involved in neonatal care in the delivery room. A date was defined for all professionals for each hospital in order to include as many staff member as possible for this study. Participation was defined as a teaching session. The participating professionals were asked to assign the AS for all 15 sequences. Altogether, 55 midwives, 68 obstetricians and 43 neonatologists from 4 general hospitals and from 3 university hospitals in Switzerland participated in the study. No sample size calculation was performed as we chose an observational approach.

Prior to scoring the study sequences, the participants were informed about the purpose of the study. A few delivery room resuscitations were shown using test video sequences in order to accustom them to perform the AS from video sequences in the same time frame as in the real delivery room situation. However, no teaching regarding the assessment of the AS was performed. All participants of a given hospital were shown the video sequences simultaneously on the same screen. No discussion was allowed among the participants. Between the sequences, short breaks were interposed to allow for noting the scores.

Statistics

Statistical analysis was performed using R (R free software environment for statistical computing and graphics, Free Software Foundation's GNU General Public License; www.r-project.org). The AS for each of the 15 patients was evaluated by midwives, obstetricians and neonatologists from seven different hospitals. Thus, each newborn infant was scored by a total of 166 observers. The objective was to estimate the variance components and to evaluate assignable causes of variability in the assigned AS. The Intra-Class Correlation Coefficient (ICC) was calculated to evaluate the inter-observer variability. Ideally, most of the variation should be explained by newborn infant's differences, and the calculated ICC should thus be high (close to 1), or low if the source of variation was due to the observers or error. Conventionally, an ICC > 0.75 is defined as high. To demonstrate if there was a significant difference in evaluating the AS depending on the different professions or on the hospital setting, a linear mixed effect model was used that incorporated both fixed and random effects.

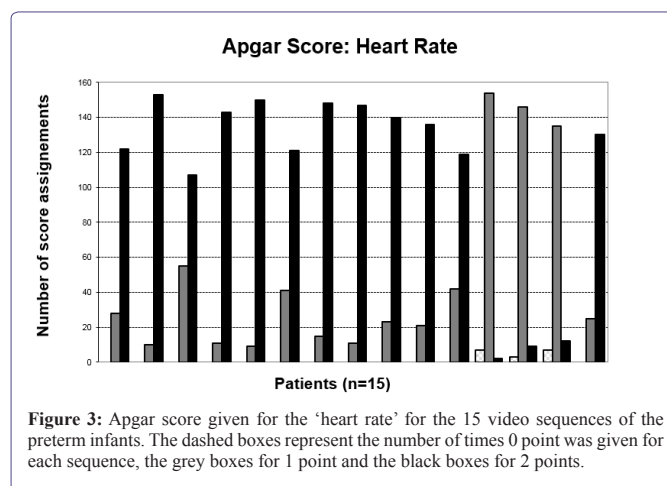
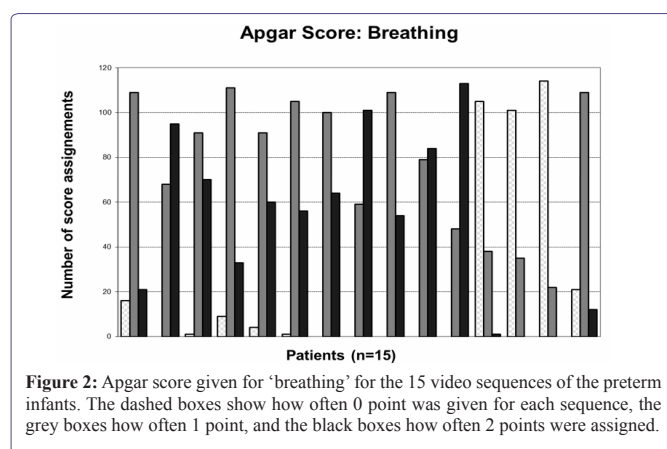
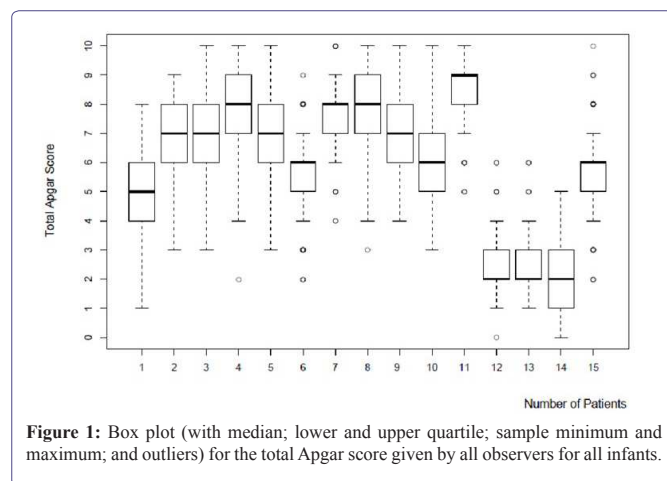
In addition, the standard deviation of the AS across the observers was calculated, yielding a standard deviation score among observers for each patient. The mean and range of these standard deviations were then computed for all patients, providing a quantitative measure in the Apgar unity as to how the observers varied in the evaluation of the AS.

To control if there were significant differences between the hospitals and professions, an F-test was performed. Furthermore, a pairwise comparison was made to link the different professions. For the p-value, a Bonferroni correction was used. Finally, to compare the university hospitals with the general hospitals, a test based on the

method of the contrasts was used. Statistical significance was assumed for p -values < 0.05 .

Results

Figure 1 depicts the distribution of the total AS for each infant given by all observers. Figure 2 represents the distribution of scores assessing 'breathing' and figure 3 the distribution for the scores assigned for 'heart rate'.



The Intra-Class Correlation Coefficient (ICC) for all infants and all observers was 0.72. The mean inter-observer variability for the AS among all observers was 1.28 (minimum 1.04; maximum 1.51). Moreover, there was a significant difference between the professions ($p < 0.001$) and hospitals ($p < 0.001$). The AS assigned by neonatologists for all infants were significantly higher than the scores given by midwives ($p = 0.001$). The median Apgar score was similar for 2/3 of the determinations made by neonatologists and midwives with a difference of only -0.26 (pairwise comparison). The scores assigned by obstetricians were the lowest for all infants; the difference from neonatologists being -0.53 (pairwise comparison). For all infants, there was no significant difference between the AS assessed by professionals working in university hospitals compared to HCPs from general hospitals ($p = 0.86$).

Discussion

With regard to the assessment of the AS, our study revealed a relatively high inter-observer agreement among all observers for the whole group of preterm infants. On the other hand, a significant difference was seen between the AS given by the different professional groups. When assessed by neonatologists and using a pairwise comparison model, the Apgar scores were significantly higher than the ones assigned by midwives, although the median Apgar score was identical for 10 out of 15 infants studied. Interestingly, the lowest scores were given by obstetricians, and again there was a trend towards lower median Apgar scores assigned by obstetricians. One reason for this difference could be the fact that obstetricians assess the AS less often in preterm infants. For this gestational age group, the AS is mainly performed by neonatologists, paediatricians and midwives. The significant difference with regard to the AS between the professionals was seen across all 7 hospitals. Our results are in accordance with earlier findings by Clark et al., who demonstrated by means of a case presentation that paediatric professionals have a significant lower variability in assigning the AS than community hospital nurses [27]. On the other hand and using three written case descriptions, Rüdiger and collaborators found that the large variations seen in the AS of VLBW infants were not affected by the degree of professional experience of the 121 paediatric professionals. They found large variations for both clinical and case scores between the centres, and units assigning low median clinical AS also had low median case scores [22]. We also showed a significant difference between scores assigned in the seven participating hospitals. Interestingly, no significant difference merged between scores assigned in university hospitals compared to those assigned in general hospitals. In a study performed by O'Donnell, ten-second video clips displaying neonatal resuscitation of 30 newborns were shown to 42 observers of six different professions in order to assign the AS [28]. In contrast to our results, they revealed a higher inter-observer variability of 0.68, even though the heart rate was reliably monitored with a pulse oximeter. Moreover, the variability in the AS assigned by observers on the base of video recordings was higher than the variability of those attending the delivery. Similarly, the scoring did not depend on the experience of the observer.

The higher inter-observer agreement seen in our study could be explained by the different statistics used. In contrast to the study of O'Donnell, we applied a linear mixed effects model incorporating the two parameters profession and clinic into the calculation. Moreover, the number of enrolled observers was higher in our study. These two differences may explain why in our study the AS differed significantly

between the professionals, and not so in the study of O'Donnell and co-workers. In accordance with their study, the time pressure during the assessment of the AS was also problematical in our study as many observers had difficulties assigning and noting the score during the break in between the video sequences. Taking into account that in real life the AS needs to be performed quickly and at well-defined points in time after birth, the original assessment protocol was followed in both studies. Of note, Virginia Apgar herself found less variation if the AS was assigned quickly [2]. In her study, the variation of the score was only 1 point between different observers and occurred mainly in the mildly depressed group.

With regard to using the AS for premature infants, this was first done by Virginia Apgar [3]. She included 70 newborns with a birth weight between 500 g and 2500 g into her study group. The score was found to measure the relative handicaps in preterm infants not without emphasizing the need for further investigations. Although being considered a relatively objective score, single parameters such as skin colour, muscle tone and reflex irritability may depend on the physiologic maturity and therefore on the gestational age. Hegyi et al., found in 1105 preterm infants with a birth weight < 2000 g that the incidence of low Apgar scores was inversely related to the birth weight and with a significant difference for gestational age [29]. In a study by Rüdiger and co-workers, the Apgar score of 1000 very low birth weight infants was evaluated from clinical charts across seven NICUs [22]. The median clinical score for all VLBW infants clearly depended on gestational age and increased with increasing gestational age.

Looking separately at the Apgar parameters determined in our study, the heart rate showed the lowest (0.2 - 0.5 points) and skin colour the highest (0.5 - 0.7 points) standard deviation for all observers. The assessment of the skin colour seems to yield the least accuracy, which makes it the weakest parameter of the AS. Besides the above-mentioned difficulties to evaluate and interpret the skin colour in preterm infants, it has also been shown that it doesn't reflect accurately the oxygenation of the infant. O'Donnell et al. using video clips reported a wide variation in the oxygenation when comparing newborn infants considered being pink by the assessors and the pre-ductal oxygen saturation values. One explanation for the highest variability seen in our study with regard to assessing the skin colour could be due to the fact that the video sequences were shown at different sites with different technical equipment, which may well have had an impact on the general colour rendering index. Besides this technical aspect and based on the discussion above, the question whether skin colour as a proxy for the oxygenation of the brain deserves its place in the AS in the future has to be frankly asked and discussed. This is especially true on the background that a quick and accurate assessment of the cerebral oxygenation in the delivery room can only be achieved by using pre-ductal oximetry, thereby giving a reliable indication for the need of supplemental oxygen but also for steering this therapy.

The high standard deviation for the muscle tone (0.2 - 0.6 points) may be explained by the fact that this parameter could not be directly assessed by the observers themselves. Instead, they had to rely on their observation of the infant's body and limbs position and movements. As mention before, again the maturity and therefore gestational age have an influence on this parameter. Due to these maturational and technical limitations, it may well be that in our study the inter-observer variability was overestimated for skin colour and muscle tone.

Conversely, the lower variation among HCPs in assessing the heart rate is reassuring when considering the pivotal role of heart rate in determining the need for changing interventions, for escalating or de-escalating resuscitation care [30].

Although many score parameters are altered by resuscitation measures [26,28,31], there is no accepted standard for reporting the Apgar score in neonates undergoing resuscitation. Clinical practice shows that same ventilated newborn infants are scored with 0 points for missing breathing effort whereas other observers will assign 2 points based on the sufficient oxygenation due to appropriate resuscitation. The same disparity applies to nasal CPAP where some centres assign 2 points for spontaneous and regular breathing while others would only give only 1 point. Bashambu and collaborators enrolled 335 neonatologists who were shown video sequences of four delivery room cases at 1, 5 and 10 minutes of life with the task to assess the AS. They found a high inter-observer agreement for respiratory efforts, grimace and muscle tone in preterm infants in the lower and higher score range, and a disagreement which was depending on the level of respiratory intervention [32]. The introduction of an expanded AS resulted in a more detailed but also complicated score and has not been shown to improve the inter-observer variability. This may be the reason for not having gained wide acceptance so far. Of note, even though the different score parameters were described more precisely in studies using written case presentations such as in the previous study as compared to the video presentations in our study, the inter-observer variability was not lower. These observations reveal an important potential for high inter-observer variability which was also addressed by The American Academy of Pediatrics emphasizing that perinatal health care providers need to be consistent in assigning the Apgar score [33].

Additionally, a source of bias could be the participant sampling method. We tried to avoid this bias by declaring the participation as an teaching session for all staff members on duty that shift. The maximal number of video sequences shown to the perinatal health care professionals was given by the time allocated by the hospitals for the teaching session (usually 45 to 60 minutes). Each video sequence needed 2-3 minutes. Besides the above-mentioned difficulties impeding on the correct assessment of the neonatal transition of an individual infant born prematurely, there is also a potential impact on the level of population studies when it comes to the prediction of neonatal mortality and long-term outcome of this patient group. Worldwide, there is a growing interest in finding suitable benchmarking indicators for international comparisons to assess differences in interventions and outcomes in order to define a quality level of neonatal care and health based on best practices [17]. Assessing the association between AS at 5 minutes of life and mortality across European countries (Euro-Peristat Project), Siddiqui et al. found a weak correlation between neonatal mortality and AS < 7 at 5 minutes. The authors concluded that the large variations seen in the distribution of AS at 5 minutes may reflect differing national scoring practices, and that without further research into standardising the coding and reporting, the AS was not suitable for evaluating the burden of neonatal mortality across countries. In their view however, the AS remains interesting on a nation-wide level as observed trends may indicate real changes within a given country.

Conclusion

In conclusion, our study revealed a relatively high inter-observer agreement in assessing the AS for premature infants among all

perinatal health care professionals for the whole group of infants. On the other hand, a significant difference was seen between the AS given by the different perinatal professional groups and between hospitals, but not between university and general hospitals.

In our view, and with respect to the physiological applicability of the actual AS to newly born premature infants and to resuscitation measures, a clearer definition and assessment method of each Apgar parameter needs to be discussed, its relevance within the AS critically evaluated and eventually implemented into future teaching models. Video sequences seem to be a suitable teaching tool for it. This may contribute to reduce the variations between professionals and hospitals, and to increase the value of this scoring within national and international databases to describe study populations for research, for benchmarking in neonatal intensive care and for comparison of outcome data.

Acknowledgement

We are indebted to the parents for agreeing to use the recordings of their newborn infants, furthermore to the midwives and medical staff at all the participating hospitals for having taken their valuable time to participate to this study.

Declaration of Interests

The authors declare that they have no conflicts of interests in this study.

Availability of Data and Materials

The data set generated and analyzed during the current study is not publicly available, since confidentiality was guaranteed towards the respondents.

Ethics Approval and Consent to Participate

No formal approval of this study was required by the Ethics Committee of the Canton Zurich as no patient data or health data of the participating HCPs was assessed. The chosen video sequences were used only with written parental permission. Care was taken to avoid any identifier of the infant and of the attending HCPs. Participation was voluntary, health care providers were asked for consent preceding the study.

Author's Contributions

SJA, HUB and JCF wrote the research protocol and the submission to the local Ethics Committee, selected the video sequences. HUB asked for parental permission. JCF and HUB supervised the project; SJA undertook the assessment of the variability in the participating hospitals, and together with MM did the statistical analysis and drafted the manuscript. JCF and HUB reviewed the manuscript; all authors read and approved the final manuscript.

References

1. Apgar V (1953) A proposal for a new method of evaluation of the newborn infant. *Curr Res Anesth Analg* 32: 260-267.
2. Apgar V, Holaday DA, James LS, Weisbrot IM, Berrien C (1958) Evaluation of the newborn infant; second report. *J Am Med Assoc* 168: 1985-1988.
3. Apgar V, James LS (1962) Further observations on the newborn scoring system. *Am J Dis Child* 104: 419-428.
4. Casey BM, McIntire DD, Leveno KJ (2001) The continuing value of the Apgar score for the assessment of newborn infants. *N Engl J Med* 344: 467-471.
5. Finster M, Wood M (2005) The Apgar score has survived the test of time. *Anesthesiology* 102: 855-857.
6. Papile LA (2001) The Apgar score in the 21st century. *N Engl J Med* 344: 519-520.
7. Pinheiro JM (2009) The Apgar cycle: A new view of a familiar scoring system. *Arch Dis Child Fetal Neonatal Ed* 94: 70-72.
8. Drage JS, Kennedy C, Schwarz BK (1964) The Apgar score as an index of neonatal mortality. A report from the collaborative study of cerebral palsy. *Obstet Gynecol* 24: 222-230.
9. Iliodromiti S, Mackay DF, Smith GC, Pell JP, Nelson SM (2014) Apgar score and the risk of cause-specific infant mortality: A population-based cohort study. *Lancet* 384: 1749-1755.
10. Li F, Wu T, Lei X, Zhang H, Mao M, et al. (2013) The Apgar score and infant mortality. *PLoS One* 8: 69072.
11. Moster D, Lie RT, Irgens LM, Bjerkedal T, Markestad T (2001) The association of Apgar score with subsequent death and cerebral palsy: A population-based study in term infants. *J Pediatr* 138: 798-803.
12. Moster D, Lie RT, Markestad T (2002) Joint association of Apgar scores and early neonatal symptoms with minor disabilities at school age. *Arch Dis Child Fetal Neonatal Ed* 86: 16-21.
13. Thorngren-Jerneck K, Herbst A (2001) Low 5-minute Apgar score: A population-based register study of 1 million term births. *Obstet Gynecol* 98: 65-70.
14. Ehrenstein V, Pedersen L, Grijsa M, Nielsen GL, Rothman KJ, et al. (2009) Association of Apgar score at five minutes with long-term neurologic disability and cognitive function in a prevalence study of danish conscripts. *BMC Pregnancy and Childbirth* 9: 14.
15. Stuart A, Otterblad Olausson P, Källen K (2011) Apgar scores at 5 minutes after birth in relation to school performance at 16 years of age. *Obstet Gynecol* 118: 201-208.
16. American Academy of Pediatrics, Committee on Fetus and Newborn; American College of Obstetricians and Gynecologists and Committee on Obstetric Practice (2006) The Apgar Score. *Pediatrics* 114: 1444-1447.
17. Siddiqui A, Cuttini M, Wood R, Velebil P, Delnord M, et al. (2017) Can the Apgar score be used for international comparisons of newborn health? *Paediatr Perinat Epidemiol* 31: 338-345.
18. Weinberger B, Anwar M, Hegyi T, Hiatt M, Koons A, et al. (2000) Antecedents and neonatal consequences of low Apgar scores in preterm newborns: A population study. *Arch Pediatr Adolesc Med* 154: 294-300.
19. Livingston J (1990) Interrater reliability of the Apgar score in term and premature infants. *Appl Nurs Res* 3: 164-165.
20. Catlin EA, Carpenter MW, Brann BS 4th, Mayfield SR, Shaul PW, et al. (1986) The Apgar score revisited: Influence of gestational age. *J Pediatr* 109: 865-868.
21. O'Donnell CP, Kamlin CO, Davis PG, Carlin JB, Morley CJ (2007) Clinical assessment of infant colour at delivery. *Arch Dis Child Fetal Neonatal Ed* 92: 465-467.
22. Rüdiger M, Küster H, Herting E, Berger A, Müller C, et al. (2009) Variations of Apgar score of very low birth weight infants in different neonatal intensive care units. *Acta Paediatr* 98: 1433-1436.

23. Marlow N (1992) Do we need an Apgar score? *Arch Dis Child* 67: 765-767.
24. Vohr BR, Wright LL, Dusick AM, Perritt R, Poole WK, et al. (2004) Center differences and outcomes of extremely low birth weight infants. *Pediatrics* 113: 781-789.
25. Gupta S, Natarajan G, Gupta D, Karnati S, Dwaihy M, et al. (2017) Variability in Apgar score assignment among clinicians: Role of a simple clarification. *Am J Perinatol* 34: 8-13.
26. Lopriore E, van Burk GF, Walther FJ, de Beaufort AJ (2004) Correct use of the Apgar score for resuscitated and intubated newborn babies: Questionnaire study. *BMJ* 329: 143-144.
27. Clark DA, Hakanson DO (1988) The inaccuracy of Apgar scoring. *J Perinatol* 8: 203-205.
28. O'Donnell CP, Kamlin CO, Davis PG, Carlin JB, Morley CJ (2006) Interobserver variability of the 5-minute Apgar score. *J Pediatr* 149: 486-489.
29. Hegyi T, Carbone T, Anwar M, Ostfeld B, Hiatt M, et al. (1998) The Apgar score and its components in the preterm infant. *Pediatrics* 101: 77-81.
30. Perlman JM (2016) Highlights of the new neonatal resuscitation program guidelines. *NeoReviews* 17: 435-446.
31. Stark CF, Gibbs RS, Freedman WL (1990) Comparison of umbilical artery pH and 5-minute Apgar score in the low-birth-weight and very-low-birth-weight infant. *Am J Obstet Gynecol* 163: 818-823.
32. Bashambu MT, Whitehead H, Hibbs AM, Martin RJ, Bhola M (2012) Evaluation of interobserver agreement of apgar scoring in preterm infants. *Pediatrics* 130: 982-987.
33. American Academy of Pediatrics Committee on Fetus and Newborn, American College of Obstetricians and Gynecologists Committee on Obstetric Practice (2015) The Apgar Score. *Obstet Gynecol* 126: 52-55.